OBJECT DETECTION AND RECOGNITION USING MACHINE LEARNING TECHNIQUES FROM IMAGES

[1]Dr Satya Prakash Singh Assistant Professor

[2]Dr Siva Suryanarayana[3] Ms. Pammi Kumari[4]Vijaya Lakshmi Assistant Professor Research Scholar Scientific Officer

[sp.singh@bitmesra.ac.in] [cssuryanarayana@bitmesra.ac.in] [pammi0126@gmail.com][vijayalakshmi_bit@yahoo.com]

1.Department of Computer Science and Engineering, Birla Institute of Technology, Mesra Ranchi, Jharkhand.

2. Department of Computer Science and Engineering, Birla Institute of Technology, Mesra Ranchi, Lalpur Campus.

3. Department of Electronics & Communication Engineering, Birla Institute of Technology, Mesra Ranchi, Jharkhand.

4. Department of Computer Science and Engineering, Birla Institute of Technology, Mesra Ranchi, Lalpur Campus.

ABSTRACT

Deep learning-based object recognition offers a quick and precise way to forecast where an object will appear in an image. The object detector automatically learns the visual features necessary for detecting tasks using the potent machine learning technology known as deep learning. The primary objective of this project is to create a system that is able to recognize objects based on a picture or a stream of images that are transmitted to it in real time from a camera or as video that has been captured in the past. The system will draw bounding boxes around the things it has detected.The system will classify the object based on the classes to which it already belongs before classifying it further. Python programming and an approach to machine learning known as the YOLO (You Only Look Once) algorithm are applied for the goal of object detection.

Keywords: Machine Learning, Scale-invariant Feature Transform, Object Detection, YOLO, Bounding Box, Rectified Linear Unit, Convolutional Neural Network.

1. INTRODUCTION

This research devotes a significant amount of space to elaborating on the fundamental ideas that underpin the Object Detection Using Machine Learning project. The project can be categorised under the artificial neural network subfield of machine learning. The principles behind machine learning make it possible for a system to learn on its own from the sum total of its experiences, unfettered by the influence of external factors. The YOLO (You Only Look Once) algorithm,

which makes use of a Convolutional Neural Network, is utilised in order to accomplish the task of detection. **(Lin et al., 2018)**. It is an Artificial Neural Network concept called a Deep Neural Network**(Shah & Kapdi, 2017)**. The biological idea of the nervous system, in which neurons serve as the network's nodes, served as the model for artificial neural networks.

The challenge of object detection has led to the application of numerous approaches. Identifying a particular thing from a picture of numerous, intricate lines and forms is called object detection**(Han et al., 2017)**. Face recognition, object tracking, picture retrieval, and automatic parking systems all involve object detection. Applications are coming in at an increasing rate. Image classification, or more accurately image retrieval, is the primary application of object detection**(Ashtari et al., 2014)**. Deep neural networks are crucial for comprehending convolution neural networks. For a better understanding of convolution neural network concepts, papers in deep neural networks are studied**(Wang et al., 2018).**The computer vision community has made significant advancements in its understanding of automatic car licence plate detection and recognition during the past 20 years. In region-based approaches,the input picture is divided into smaller portions, and the ensuing areas are where particular elements of the licence plate are located. In**(H. Xie et al., 2015)** a pixel-to-pixel technique, each pixel in the image is compared to its nearby pixels to create a rough rectangular box.

Using a detection window, the entirety of the image that is being read in is scanned pixel by pixel. At each individual pixel location, the response of the scanning window is measured, and the areas that produce the most accurate results are identified are chosen as candidate regions. In order to classify an image pixel-by-pixel, **L. Dlagnekov (200**4) developed the Adaboost classifier. The detection of licence plates was addressed using a pixel-to-pixel approach in. The authors made a contribution by creating an Adaboost classifier initially, then a SIFT-based SVM classifier afterwards. F.A. Silva also used a Scale-invariant Feature Transform (SIFT) based technique for identifying and detecting licence plates**(Hu et al., 2016).** A fresh color-based method was subsequently developed by transforming RGB-color images to HSV-color space. Perceptrons function similarly to network nodes in artificial neural networks**(Kido et al., 2018)**. An artificial neural network will have an input layer, a hidden layer, and an output layer. These will be the three layers that make up the network. Deep learning refers to the portion of an artificial neural network that includes multiple hidden layers and is capable of being used for

feature extraction and classification. This section of the network is known as the deep learning layer.

The Convolutional Neural Network (CNN) offered by Deep Learning is a piece of software that can analyse visual data**(Abbas & Singh, 2018)**. Convolutional, pooling, activation, and fully connected layers make up its four different types of layers. Strides and a filter are used by the convolution layer to create the feature maps. Deep convolutional neural networks, or CNNs, have been utilised throughout the course of the past several years to address a wide range of issues pertaining to computer vision, and many academics have experimented with utilising CNNs to detect and recognise vehicle licence plates**(L. Xie et al., 2018).**

Research on computer vision tasks for the transportation system has advanced significantly in recent years. Some new methods have a tendency to use the characteristics recovered by CNN rather than hand-crafted features when robustness is highly demanded**(Buric et al., 2018)**. Additionally, the dataset must be enhanced with specific data augmentation techniques. The accuracy has undoubtedly increased as a result of the effort done to balance the proportion of various objects in the dataset using hierarchical augmentation(**Han et al., 2017)**. Additionally, a lot of studies have explored using prior knowledge, like location or symmetry, to enhance performance**(Bisong, 2019)**. Because of their high level of generalisation and the fact that they can be used to construct non-linear classifiers, Support Vector Machines (SVM) have gained popularity ever since its inception as a classification method.

It has been observed that these classifiers have been successfully applied in a number of different ways for detecting and tracking. SVM classifiers are extremely fascinating when applied to the problem of intelligence in video surveillance, for example, because of the possibility to zero in on a specific class of natural objects that may be learnt from training data. This makes SVM classifiers quite appealing**(Han et al., 2017)**. Therefore, the ability to learn statistically variable natural objects is necessary for the detection of specific objects in video analysis. Examples of such items are the tail ends of cars and faces**(Carminati et al., 2006)**.

The capacity to detect and follow moving objects on the basis of their membership in the same preset class, as opposed to the fact that they are merely moving, is an aspect that is even more fascinating. In real-time applications, SVM-based identification and tracking do have some limitations, however. In point of fact, when working in pixel space, object classification is quite

laborious, and feature selection approaches like contour or colour extraction are unable to be adequately discriminative**(Budiharto et al., 2018)**. The problem is made significantly more difficult for the tracking procedure when objects' motion is taken into account. The matrix that is obtained following the convolution layer is these feature maps. ReLU (Rectified Linear Unit), which converts negative numbers to 0, helps simplify things. By passing the generated Feature Map into the Pooling Layer, it is shrunk to a smaller sized matrix. The features are extracted in this manner. The Fully Connected Layer, located at the convolutional neural network's conclusion, is where the classification process actually takes place.

2. MACHINE LEARNING FUNDAMENTALS

Classifier based on the logistic regression and the softmax function

The logistic classifier is the most typical example of a linear function, and it is utilised extensively in the process of data training. The logistic classifier is made up of weights and biases, both of which are employed in order to tune the variations in the data. The category of the point is determined by its weight**(Abbas & Singh, 2018)**.

$$Wx + B = y$$

Where W = weights, x= inputs, B=biases, y=output

The equation takes a vector as its input, and its output is a one-dimensional array of values, with the greatest value corresponding to the most likely label. The vector serves as the equation's input. This output is then converted into the probabilities by the softmax function, which works under the assumption that the sum of all probabilities equals one and that the item with the greatest output value has the highest likelihood.

After that, the output of the softmax is turned into 1 hot encoding vectors, which tell the user which class label is being predicted. The entirety of the map can be seen in Figure 1**(Abbas & Singh, 2018)**.

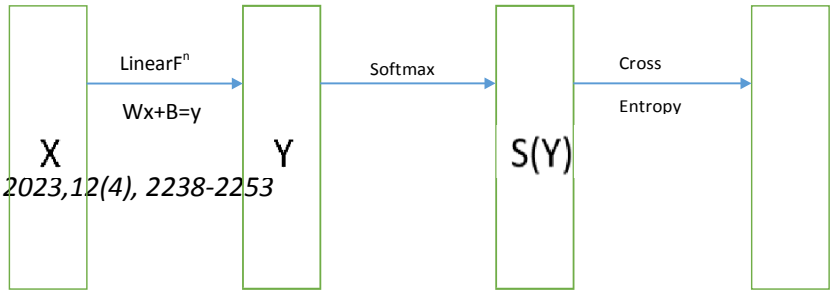

Fig1: Multinomiallogisticclassification

3. DEEP NEURAL NETWORKS

The artificial neural networks have been extended in scope to become the deep neural networks. The term "deep neural network" refers to a network whose depth can be increased by increasing the depth of an artificial neural network. Networks are considered to be deep if their depth is greater than three or if they have more than three layers. The ability to utilise non-linearity is the primary benefit that comes from making use of these strategies. In order to make use of nonlinearity, Rectified Linear Units (RELU) can be implemented in the hidden layers of the model. There are a total of (n+1) x k parameters in a deep neural network, where n represents the input and k represents the output.

When compared to employing a non-linear function, using a linear function has a number of distinct advantages. A linear function will always produce the same result when differentiated. When applied to a linear function, GPU matric multiplication is significantly faster than when applied to a non-linear function. The output of linear functions is very steady, meaning that even minute changes can have a noticeable impact. Therefore, a linear function may be utilised during the implementation of a neural network with multiple layers.

A. RELU (Rectified Linear Units)

RELU functions are the functions which convert non-linear equation into a linear one. The RELU function is defined as

$$y = \begin{cases} 0 \; if \; x < 0 \\ x \; if \; x \geq 0 \end{cases}$$

One of the benefits of this function is that its derivative is a constant function. This is an advantage over other functions. Therefore, the RELU function is employed whenever an additional hidden layer is added to a neural network.

4. STATEMENT OF THE PROBLEM

The "Object Detection System using Machine Learning Technique from Images" project successfully detects objects using the YOLO algorithm and applies the algorithm to both image and video data. The study shall provide valuable insight and knowledge about the effect of Object Detection System using Machine Learning Technique from Images.

5. THE PROPOSED MODEL'S ARCHITECTURE

The proposed YOLO model's architecture diagram is shown in Fig. 1. Images are provided to the system as input. Since video is only a stream of images, it may also be used as an input. As the name implies, You Only Look Once only passes the input is run through the network just once to get the result, which is the location of an object along with bounding boxes and labels.

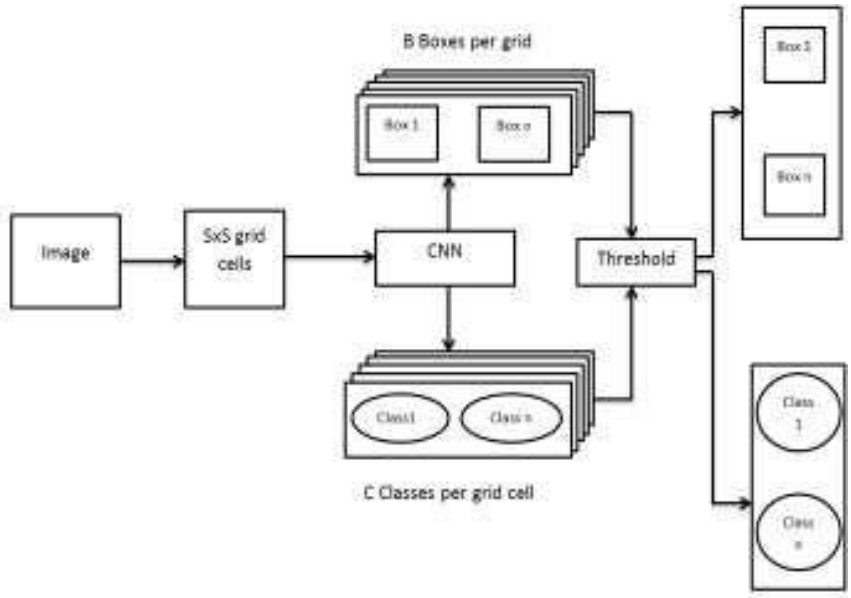

Fig 2: YOLO Architecture

Before submitting the images to the Convolutional Neural Network, they are split into SXS grid cells (CNN). B The Convolutional Neural Network produces bounding boxes for each grid around each recognised object in the image. On the other side, the Convolutional Neural Network also classifies the Classes to which the objects belong, producing C Classes per grid. The Object Detection is then given a threshold. We set the Threshold for this project at

0.3.Lowering the Threshold value will cause the output to become clunky as more bounding boxes will be shown.

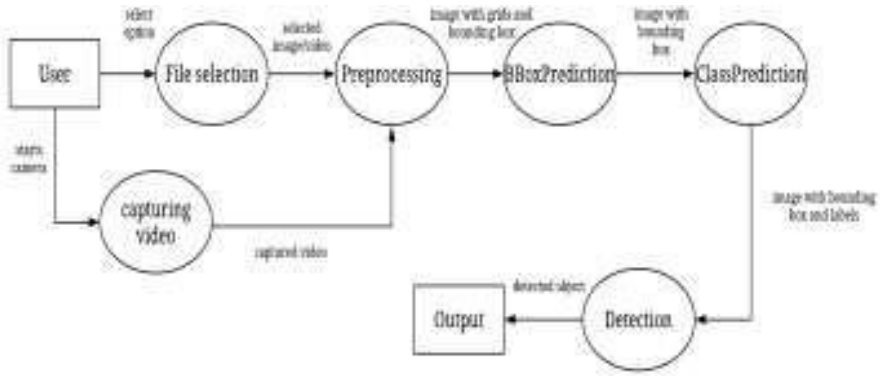

Fig 3: Data Flow Diagram of the System

The data flow in the system is shown in Fig. 2. The user will initially have the ability to select the type of file to input into the system. Therefore, the user has two options: start the camera or select a file. The user can choose an image file or a video file in the first option, and the camera module can be launched in the second. After choosing the input, preprocessing is completed, and SXS grids are created. Grids are utilised to generate the result, which is then passed on to the Bounding Box Prediction algorithm. This is the step at which bounding boxes are created around the things that have been identified. After receiving the data from the phase before it, the Class Prediction method then makes an educated guess as to the Class to which the item in question belongs. After that, it is passed on to the detecting phase, during which a Threshold is established in order to reduce the clumsiness of the output as well as the various bounding boxes and labels. An image or stream of images is produced at the completion of the process when an image, video, or camera is used as the input, and the output consists of bounding boxes and labels.

5. IMPLEMENTATION

In this research, the methodology that will be used to implement this project is described. The procedure for locating the object using the Object Detection System is outlined in the following algorithm.

A Method for the Object Detection System's Algorithm

1. An SxS grid is used to divide the input image.

2. It forecasts B boundary boxes for each cell. Five components make up each bounding box: (x, y, w, h) and a box confidence score.

3. No matter how many bounding boxes are present, YOLO only recognises one item per grid cell.

4. It makes C conditional class probabilities predictions.

5. If there are no objects, the confidence score is zero. Otherwise, the confidence score must be higher than or equal to the threshold value.

6. YOLO then creates a bounding box around the objects it has spotted and predicts which class they belong in.

6. RESULTS AND ANALYSIS

This paper explains the system's outcomes as well as the many Test Cases that were employed. We used the COCO pretrained dataset, which contained 80 classes. The data would have been overfitted with more classes, which is why there are only 80 of them. The following part will outline the various Test Cases and the outcomes.

Test Cases

The many Test Cases, along with both the Expected and the Test Results, are presented in Table 1.

Table -1: Test Cases with Results

Test CaseID	TestConditions	ExpectedResult	TestResults
TC1	When imageis chosen asinput	Image withbounding boxaround theobjects andpredictedclass	SUCCESSFUL

TC2	Whenvideoisc hosen asinput	Video withbounding boxaround theobjects andpredictedclass	SUCCESSFUL
TC3	When camerais chosen asinput	Objectsdetected in thereal time withbounding box,confidencescore andpredictedclass	SUCCESSFUL
TC4	When blackand whiteimageist akenasinput	Image withbounding boxaround theobjects andpredictedclass	SUCCESSFUL
TC5	Imagewithfar objects aretakenasinp ut	Imagewith detectedobjects	UNSUCCESSFUL
TC6	When imagewith overlappingo bjectsis taken as input	Image withboundingbox around theobjectsand predicted class	SUCCESSFUL
TC7	When image with far objects is taken as input	Image with detected objects	UNSUCCESSFUL

Results

In this section, the various findings that were achieved by delivering the various Test Cases indicated above are described.

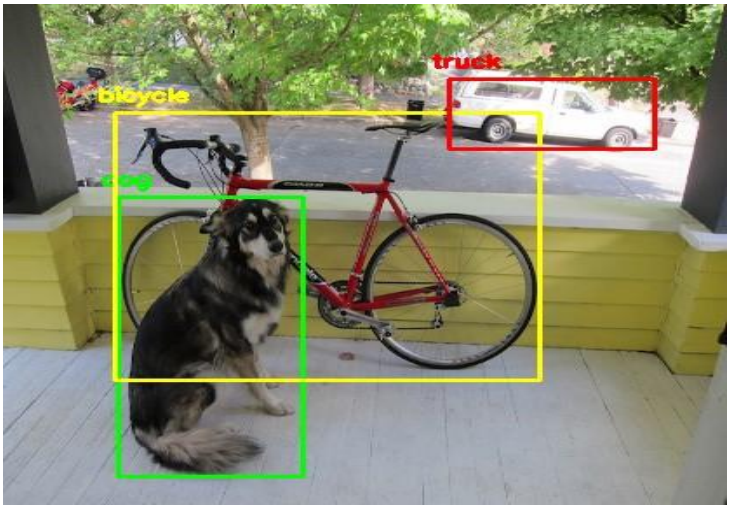

Fig-4: The result of the Object Detection System is depicted here by an image including the detected object. The objects that are detected have bounding boxes drawn around them.

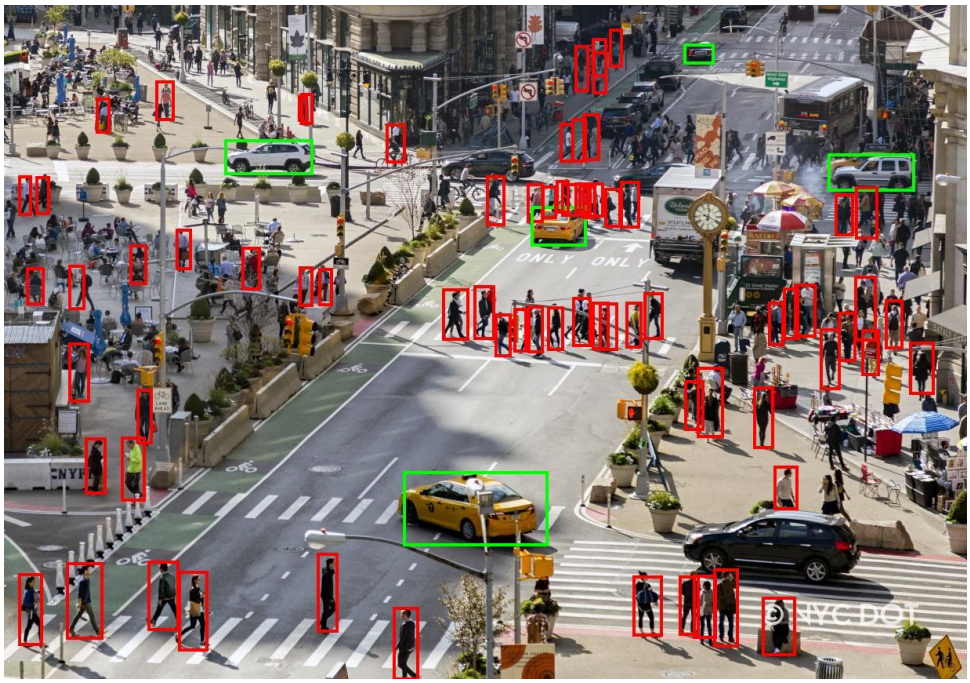

Fig-5:ImagewithOverlappingObjects illustrates the output obtained when

objects are overlapping. This shows that partially visible objects will also be detected by drawing bounding box around it along with the label some people are partially visible in the image of a crowded place. The system is able todetect every person visible in the image.

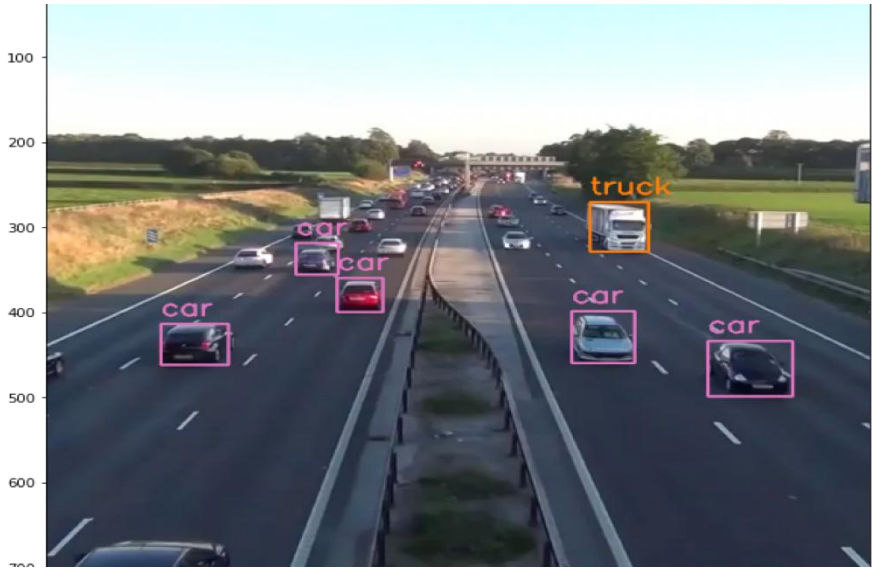

Fig-6:OutputobtainedwithVideoInput illustrates that is to be given as input to the system should be in .avi format.

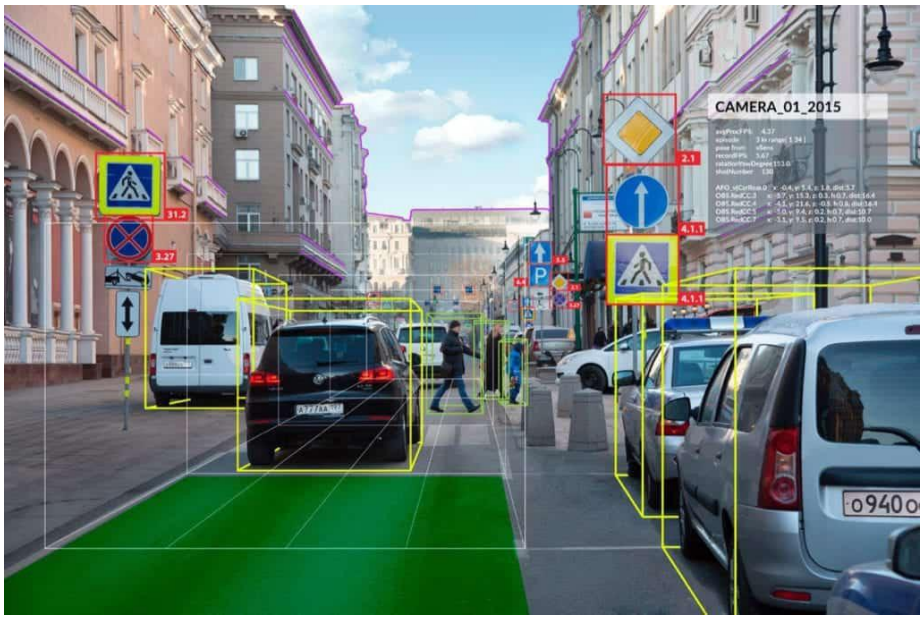

Fig -7: Output obtained in Real-Time illustrates the outpu twhen Camera is used to detect the object.

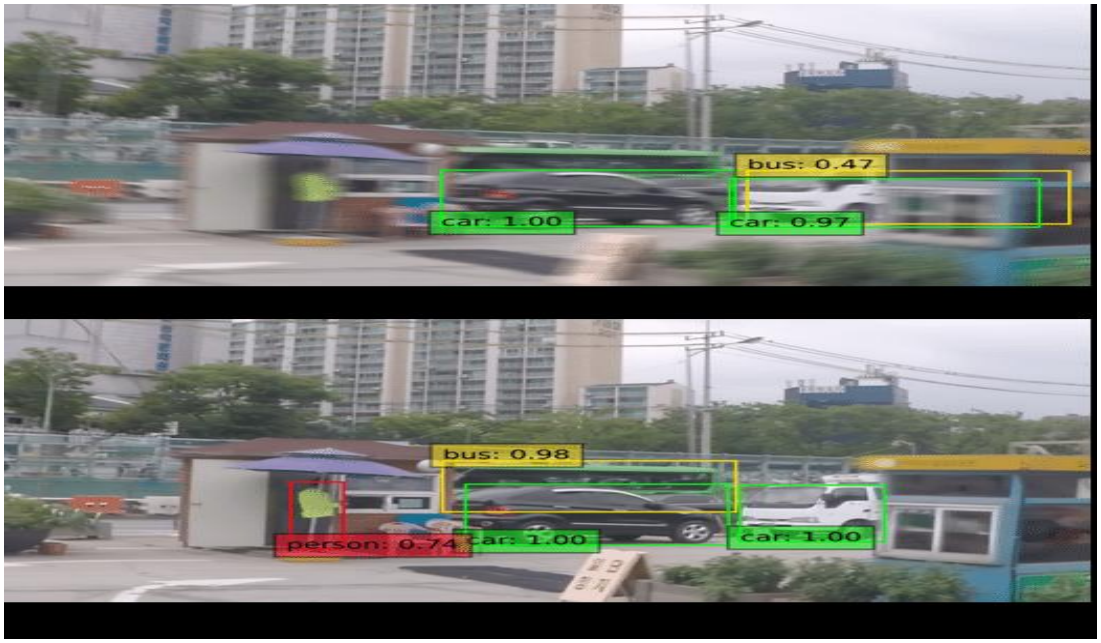

Fig -8: Output obtained in Real-Time illustrates the output generated when a blur image is given as the input. Even if no item is recognised, a set of bounding boxes will be drawn at random. This is an issue that has arisen during testing and has led to negative outcomes forthe project.

7. CONCLUSION

The performance of a model for image classification is evaluated using the mean classification error across all of the projected class labels. The objective of the project is to recognise things in real time within photos, videos, and camera feeds. The performance of a model for single-object localization is evaluated by comparing the expected bounding box to the predicted bounding box for the expected class and finding the distance between the two. Whereas the performance of a model for object identification is evaluated by applying the accuracy and recall across each of the best matching bounding boxes for the known objects found in the image. Now that we understand the challenge of object localization and detection, we can go on. This study shown, as a conclusion, that bounding boxes are drawn around the items that have been detected, and within each box is a label that indicates the class to which the object belongs. We made use of a CPU so that the project could be processed.

8. FURTHER RECOMMENDATIONS

- Repeating the study with a bigger sample size and over a longer time frame is recommended.
- The development of computer vision systems has focused a lot on efficient and precise object detection. Deep learning approaches have significantly improved the treatment for object detection.
- The project can be executed on various system for quicker results and greater precision, concentrating future developments.

REFERENCES

Abbas, S. M., & Singh, D. S. N. (2018). Region-based Object Detection and Classification using Faster R-CNN. *2018 4th International Conference on Computational Intelligence & Communication Technology (CICT)*, 1–6. https://doi.org/10.1109/CIACT.2018.8480413

Ashtari, A. H., Nordin, M. J., & Fathy, M. (2014). An Iranian License Plate Recognition System Based on Color Features. *IEEE Transactions on Intelligent Transportation Systems*, *15*(4), 1690–1705. https://doi.org/10.1109/TITS.2014.2304515

Bisong, E. (2019). Regularization for Deep Learning. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform* (pp. 415–421). Apress. https://doi.org/10.1007/978-1-4842-4470-8_34

Budiharto, W., Gunawan, A. A. S., Suroso, J. S., Chowanda, A., Patrik, A., & Utama, G. (2018). Fast Object Detection for Quadcopter Drone Using Deep Learning. *2018 3rd International Conference on Computer and Communication Systems (ICCCS)*, 192–195. https://doi.org/10.1109/CCOMS.2018.8463284

Buric, M., Pobar, M., & Ivasic-Kos, M. (2018). Object detection in sports videos. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1034–1039. https://doi.org/10.23919/MIPRO.2018.8400189

Carminati, L., Benois-Pineau, J., & Jennewein, C. (2006). Knowledge-Based Supervised Learning Methods in a Classical Problem of Video Object Tracking. *2006 International Conference on Image Processing*, 2385–2388. https://doi.org/10.1109/ICIP.2006.312942

Han, K., Sun, M., Zhou, X., Zhang, G., Dang, H., & Liu, Z. (2017). A new method in wheel hub surface defect detection: Object detection algorithm based on deep learning. *2017 International Conference on Advanced Mechatronic Systems (ICAMechS)*, 335–338. https://doi.org/10.1109/ICAMechS.2017.8316494

Hu, Q., Paisitkriangkrai, S., Shen, C., van den Hengel, A., & Porikli, F. (2016). Fast Detection of Multiple Objects in Traffic Scenes With a Common Detection Framework. *IEEE Transactions on Intelligent Transportation Systems*, *17*(4), 1002–1014. https://doi.org/10.1109/TITS.2015.2496795

Kido, S., Hirano, Y., & Hashimoto, N. (2018). Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN). *2018 International Workshop on Advanced Image Technology (IWAIT)*, 1–4. https://doi.org/10.1109/IWAIT.2018.8369798

Lin, Y.-L., Chiang, Y.-M., & Hsu, H.-C. (2018). Capacitor Detection in PCB Using YOLO Algorithm. *2018 International Conference on System Science and Engineering (ICSSE)*, 1–4. https://doi.org/10.1109/ICSSE.2018.8520170

Shah, M., & Kapdi, R. (2017). Object detection using deep neural networks. *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 787–790. https://doi.org/10.1109/ICCONS.2017.8250570

Wang, J., Wang, W., & Gao, W. (2018). Multiscale Deep Alternative Neural Network for Large-Scale Video Classification. *IEEE Transactions on Multimedia*, *20*(10), 2578–2592. https://doi.org/10.1109/TMM.2018.2855081

Xie, H., Wu, Q., Chen, B., Chen, Y., & Hong, S. (2015). Vehicle Detection in Open Parks Using a Convolutional Neural Network. *2015 Sixth International Conference on Intelligent Systems Design and Engineering Applications (ISDEA)*, 927–930. https://doi.org/10.1109/ISDEA.2015.233

Xie, L., Ahmad, T., Jin, L., Liu, Y., & Zhang, S. (2018). A New CNN-Based Method for Multi-Directional Car License Plate Detection. *IEEE Transactions on Intelligent Transportation Systems*, *19*(2), 507–517. https://doi.org/10.1109/TITS.2017.2784093

Authors

1. **Dr. Satya Prakash Singh** is presently working as Assistant Professor (Selection Grade) in Department of Computer Science & Engineering, Birla Institute of Technology, Mesra, Ranchi, India. He has 24 years of teaching experience. He has received his Ph. D. degree in Computer Science and Engineering in 2012. He has received M. Tech. degree in Computer Science & Engineering in 1999. He has received M. Sc. Physics (Electronics) degree in 1997. He has supervised many research scholars of MCA, M. Tech. and Ph. D degrees. He has published more than 65 research papers in Scopus indexed journals; IEEE, Springer and Elsevier conferences, and many other national and international journals and conferences. He has been editorial and reviewer in Journal of Green Computing(Index in Scopus). His areas of interests include artificial intelligence, Data Mining, Image Processing. Parallel and Distributed Computing, and Software Engineering.

2. Dr. Siva Suryanarayana Chintapalli. is presently working as an **Assistant Professor** in Computer Science and Engineering Department at Birla Institute of Technology, Mesra, Ranchi(Lalpur Campus), India. He has received Ph.D. in Information Technology degree in 2021 from Birla Institute of Technology, Mesra, Ranchi, India.He has received M. Tech (Information Technology) degree in 2005 from AAI, Allahabad, Uttar Pradesh, India. He has received M.C.A. degree in 1996 from Osmania University, Hyderabad, Telangana, India.

He has 25 years of Teaching experience. He has supervised many scholars of MCA and B. Tech. degrees. He has published 08 research papers.

3. Ms. Pammi Kumari, Research Scholar, Department of Electronics & Communication Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand 835215, India. She is doing Ph. D. She has received B.Tech, and M.Tech degree in of Electronics & Communication Engineering. She has 7 years teaching and research experience in the same area. She has published many papers in the International, National conferences and reputed Journals. Her research interests include Medical Image Processing, Artificial intelligence, and Machine Learning.

1. Mrs. Vijaya Lakshmi Chintapalliis presently working as an Jr. Scientific Officer in Computer Science and Engineering Department at Birla Institute of Technology, Mesra, Ranchi, Lalpur Campus. She has received MPhil. in Mathematics in 2012. She received her M. Sc in Mathematics from Nagarjuna University.